Forced evolution *in silico* by artificial transposons and their genetic operators: The ant navigation problem

Leonid Zamdborg [a], David M. Holloway [b], Juan J. Merelo [c], Vladimir F. Levchenko [d], Alexander V. Spirov [a,d,e,*]

[a] *Department of Applied Mathematics and Statistics, The Stony Brook State University, NY, United States*
[b] *Mathematics Department, British Columbia Institute of Technology, 3700 Willingdon Avenue, Burnaby, BC V5G 3H2, Canada*
[c] *Departamento de Arquitectura y Tecnología de Computadores, University of Granada, Granada, Spain*
[d] *The Sechenov Institute of Evolutionary Physiology and Biochemistry, 44 Thorez Ave., St. Petersburg 194223, Russia*
[e] *Computer Science Department and Center of Excellence in Wireless & Information Technology, The Stony Brook State University, NY, United States*

ARTICLE INFO

ABSTRACT

Modern evolutionary computation utilizes heuristic optimizations based upon concepts borrowed from the Darwinian theory of natural selection. Their demonstrated efficacy has reawakened an interest in other aspects of contemporary biology as an inspiration for new algorithms. However, among the many excellent candidates for study, contemporary models of biological macroevolution attract special attention. We believe that a vital direction in this field must be algorithms that model the activity of "genomic parasites", such as transposons, in biological evolution. Many evolutionary biologists posit that it is the co-evolution of populations with their genomic parasites that permits the high efficiency of evolutionary searches found in the living world. This publication is our first step in the direction of developing a minimal assortment of algorithms that simulate the role of genomic parasites. Specifically, we started in the domain of genetic algorithms (GA) and selected the artificial ant problem as a test case. This navigation problem is widely known as a classical benchmark test and possesses a large body of literature. We add new objects to the standard toolkit of GA – artificial transposons and a collection of operators that operate on them. We define these artificial transposons as a fragment of an ant's code with properties that cause it to stand apart from the rest. The minimal set of operators for transposons is a transposon mutation operator, and a transposon reproduction operator that causes a transposon to multiply within the population of hosts. An analysis of the population dynamics of transposons within the course of ant evolution showed that transposons are involved in the processes of propagation and selection of blocks of ant navigation programs. During this time, the speed of evolutionary search increases significantly. We concluded that artificial transposons, analogous to real transposons, are truly capable of acting as intelligent mutators that adapt in response to an evolutionary problem in the course of co-evolution with their hosts.

© 2015 Elsevier Inc. All rights reserved.

* Corresponding author at: State University of New York at Stony Brook, Computer Science Department and Center of Excellence in Wireless & Information Technology, Stony Brook University Research & Development Park, 1500 Stony Brook Road, Stony Brook, NY 11794-6040, United States. Tel.: +1 (631) 632 5582; fax: +1 (631) 632 4653.
E-mail address: Alexander.Spirov@gmail.com (A.V. Spirov).

1. Introduction

Many areas of Evolutionary Computation (EC), especially genetic algorithms (GA) and genetic programming (GP), were inspired by ideas from evolutionary biology. However, modern evolutionary biology has since advanced considerably, revealing that Darwinian evolution (microevolution) is apparently a particular case of the greater mechanisms of macroevolution. Many current branches of research in evolutionary computation implement the evolution of mechanisms, such as functional programs, neural networks, decision trees, cellular automata, L-systems, and finite state automata. The recent achievements in genomics seem more appropriate as an inspirational model for these domains than the classical set of Darwinian algorithms [Cf. 2,7,13,33,37,40].

This stimulates us to select, formalize and apply to EC new evolutionary mechanisms that may help to simulate the creative, heuristic and self-organizing character of biological evolution [64–76].

We concentrate on the mechanisms of natural genetic engineering, whose key actors are mobile genetic elements (MGE) [Cf. 40,41,56,57]. The basic idea can be outlined as follows:

- The genome of every organism has mechanisms for genomic rearrangement.
- These mechanisms are activated during periods of evolutionary crisis.
- The mechanisms cause multiple systemic rearrangements of genomes within a few generations.

The key players of natural genetic engineering are mobile genetic elements (synonymous or related terms are jumping genes, selfish DNA, transposons and retroelements) [21,38,41,42]. Many biologists speculate that processes in the world of transposons, existing on a substratum of genomes of the greater biological community, are the main source of evolutionary creativity [10,26,56,57].

It has been estimated that the majority of the DNA in higher organisms is neither translated into proteins nor involved in gene regulation, and is simply "junk" DNA. The bulk of junk DNA is composed of intermediate-repeats comprised of DNA elements that are able to move (or transpose) throughout the genome. These mobile DNA elements are sometimes termed "selfish", since they do not appear to directly benefit the host. Their behavior looks parasitic: they jump between different sections of a genome in order to propagate themselves, a behavior that seems at first glance to be usually detrimental to their host. However, they do appear to benefit the host organism by providing genomic rearrangements that permit the evolution of new genetic networks.

There are several groups of genomic parasites, with transposons forming the most sophisticated one. A transposon is a mobile piece of DNA that is flanked by terminal repeat sequences and typically bears genes coding for transposition functions. "…it is now recognized that a significant portion of the genome of any eukaryote is composed of "selfish" or "parasitic" genetic elements, which gain a transmission advantage relative to other components of an individual's genome, but are either neutral or detrimental to the organism's fitness" [21]. Arguments have been made that the structure of eukaryotic genomes, including the abundance of transposons, repetitive DNA and introns, provide high evolvability despite possible detrimental effects on the host [56,57].

Transposons are ubiquitous and may comprise up to 45% of an organism's genome [21,38,41,42]. DNA sequences of transposon origin can be recognized by their palindrome endings, flanked by short, non-reversed, repeated sequences resulting from insertion after staggered cuts. Many transposons have a unique DNA site that acts as a forwarding address, directing the transposon to a complementary site elsewhere in the host genome. There are usually multiple copies of any given DNA site in the host genome, and the transposon will attach to a proper site in a random manner. Transposons may grow by acquiring more sequences; one such mechanism involves the placement of two transposons into close proximity so that they act as a single large transposon incorporating the intervening code.

During the last two decades, results were obtained that give credence to the idea that transposable elements are a major source of genetic change, such as the creation of novel genes, the alteration of gene functions, and the genesis of major genomic rearrangements [16,17,21,38,41,42,50,80]. The long coexistence of transposable elements in the genome is expected to be accompanied by host – transposon co-evolution. There exists an opinion that transposons should be treated as non-random (or rather "intelligent") mutators. Furthermore, transposons can be considered an evolutionary "SOS-crew", cooperatively acting "in emergency", during periods in which the host is experiencing genomic stress [38,44].

Modern evolutionary biology and evolutionary genetics (and genomics) amassed a great deal of knowledge regarding the complex mechanisms of genomic rearrangement governed by transposons [1,10,21,26,41,42,56,57]. From an EC point of view, the most sophisticated of these mechanisms may be treated as a local evolutionary search. By this, we mean the mutational hot spots generated by transposons. A promising approach is to include the world of selfish elements in the standard toolkit of EC, and to use the objects/procedures for maintaining and manipulating transposons as a new branch of EC. On the other hand, as was pointed in [7], EC "would enable simulations to explore how endosymbiosis or independently replicating transposable elements might affect the complexity of the genome."

Several teams have applied algorithms inspired by jumping genes (summarized in the next section). However, these have been mainly tested on benchmarks with binary or real number representations, and mainly on commonly known mathematical test-functions. In this article, we apply modern evolutionary ideas to well-known benchmark tests implemented (implicitly or explicitly) via symbolic strings. In particular, we propose an improvement on the animate pathfinder (artificial